# The Use of Hierarchical Cluster Analysis in Classifying the Phenomenon of Drug Addiction in the Governorates of Iraq, Except for the Kurdistan Region

**Sarah Ismael Khalel**
Department of Mathematics
College of Education for Girls
Sarah87omar@gmail.com

**Dr. Ammar Kuti Nasser**
Department of Mathematics
College of Basic Education
Dr.ammar168.edbs@uomustansiriyah.edu.iq

**Nihad Sharif**
Department of Mathematics
College of Education for Girls
Nihad.shreef16@tu.edu.iq.

**Abstract**

In this research, hierarchical cluster analysis was used, which is one of the types of multivariate cluster analysis. A group of governorates, except for the Kurdistan region of Iraq, were classified into homogeneous groups for drug addicts in terms of gender, age and occupation. The segmentation method was used to consist of four clusters, each cluster consisting of a number of homogeneous observations among them. Therefore, the observations of the first group should be more homogeneous than the observations of the second group.

**Keywords:** Single Linkage, Complete Linkage, Average Linkage, Multiviriat Analysis.

## 1.1: Introduction:.

The hierarchical cluster analysis is one of the most important methods in cluster analysis because it depends on simple foundations and works on the vocabulary of the sample (n) single and consecutively with (m) a cluster. Because it contains all the vocabulary of the sample [4] .

And that each cluster consists of many close groups that are linked to each other by means of relationships that fulfill the preferred conditions of convergence .

We have previously shown that the hierarchical cluster analysis classifies the sample items using one of the following two methods:

**The First topic**

### 1.1.1: Agglomerative method:

It starts from the vocabulary, and each of them constitutes an independent cluster. Then it works on merging the most similar vocabulary and forming from them similar clusters and then merging similar clusters into larger ones. Until a single cluster is obtained that includes all the items of the sample [5].

### 1.1.2: Divisive method:

This method considers that all the vocabulary forms one cluster. And then it is divided into similar clusters and then it is divided into clusters smaller than that in order to get clusters for each item of the sample items or stop the division process

And the results we get from these two methods represent a level data in the form of interconnected clusters within one cluster [3]

**1.2: Methods of aggregative hierarchical cluster analysis:**

**1.2.1: Single Linkage Method:**

This method starts from considering each of the vocabulary of forming a special cluster. It depends on the matrix of distances D or the similarity S between the pairs of items studied. Thus, the clusters are formed by merging the most closely related clusters (nearest neighbourhood).

To determine the two clusters that are more convergent, we study the elements of matrix D' and determine the smallest element in it. And suppose that it corresponds to the two clusters u, v, so we work by merging these two clusters into a new cluster and denoting it with the symbol $(u, v)$ and to calculate the radius of the new cluster distances $(u, v)$ we replace every two opposite elements of $(v)$ and $(u)$ with its smallest (nearest) We apply the following relationship:

$$d_{(u,v)} = \min_{j \in u\,,\, k \in v} [d_{jk}] \dots\dots (1)$$

where j belongs to u and k belongs to $v$ .

To calculate the ray distances between the new cluster $(u, v)$ and any other cluster or (single) $W$, we apply the following relationship

$$d_{(u,v)w} = min[d_{uw}\,, d_{vw}] \dots\dots\dots (2)$$

So, $dvw$ and $duw$ are the distances between the most closely related cluster u and the $w$ cluster. The cluster $v$ is closest to the cluster $w$, respectively [1].

**1.2.2: Complete Linkage:**

This method also starts from considering each of the vocabulary forms a special cluster. It depends on the matrix of distances $D$ between the studied vocabulary pairs. Also, the process of forming clusters in it can be done by merging the more closely related clusters in order to maintain the similarity internally within the new clusters, but the process of calculating the elements of the ray distances for the new cluster. This is done by replacing every two opposite elements of the two merged clusters with the largest (to the next to the farthest), so the computation process starts by searching for the smallest element in the distance matrix $D$. Then the two most closely related clusters are determined. We symbolize them with the symbol $u, v$ and then we combine these two clusters into a new cluster symbolized by the symbol $(u, v)$ and to calculate the ray distances for the new cluster $(u, v)$ We replace every two opposite elements in u, v with the largest (with the farthest), that is, we apply the following relationship [6] :

$$d_{(u,v)} = \max_{j \in u\,,k \in v} [d_{jk}] \dots\dots (1)$$

As: $j \in u\,, k \in v$

Then we calculate the ray distance between the new cluster $(u, v)$ and any cluster or individual W from the following relationship:

$$\dots\dots\dots (2) d_{(u,v)w} = \max [d_{uw}\,, d_{vw}]$$

When performing the calculations, we follow the same procedures that we followed in the method of single linking, with the only difference being the method of calculating the

new distances. Since the distances of the new cluster $d$ $((u,v)$ ) are calculated by taking the largest of any two opposite elements of $v,u$. The above two relationships can be applied to the trigonometric matrix $D$. The comparison method can also be applied to the basic matrix, as we did in the case of the single link [6].

**1.2.3: Medium method:**

This method also starts from considering each of the items as a special cluster, and depends on the matrix of distances $D$ between the pairs of items studied. Clusters are formed by merging the most closely related clusters. But the process of calculating the ray distance elements of the new cluster is done by taking the arithmetic mean of each two opposite elements of the two merged clusters.

To apply this method, we first study the elements of the matrix $D$ and determine the smallest in it, and suppose that it corresponds to the two clusters $v,u$ in a new cluster and denotes it with $(u,v)$. And we put them in a column and a line dedicated to the new cluster $(u,v)$, that is, we calculate the average distances of the cluster $(u,v)$ from the relationship:

$$d_{(u,v)j} = \frac{1}{2}\left[d_{uj} + d_{vj}\right] \dots\dots\dots\dots (1)$$

The ray distance between the new cluster $(u,v)$ and any other cluster, $w$ or singular, can also be calculated from the relationship:

$$d_{(u,v)} = \frac{\sum_{j=1}\sum_{k=1} d_{jk}}{n_{(uv)} * n_w} \dots\dots\dots\dots . (2)$$

Since: $n_{uv}$ is the number of items in the cluster $(u,v)$

that : $n_w$ is the number of items in the cluster $(w)$

that : $d_{jk}$ is the distance between the j element of $(u\ ,v)$ and the k element of $(w)$

There are other formulas for calculating these distances, the most important of which is the following (ward) formula [web p. 368]

$$d_{(i+j)k} = \frac{n_k + n_i}{n_k + n_i + n_j} d_{ik} + \frac{n_k + n_j}{n_k + n_i + n_j} d_{jk} - \frac{n_k}{n_k + n_i + n_j} d_{ij} \dots\dots . . (3)$$

So: $n_i, n_j, n_k$ is the number of items in the groups $(i,j,k)$ respectively.

As each of $d_{ik}, d_{jk}$ is the Euclidean distance between the two terms $(i,k)$, $(j,k)$ respectively.

The $d_{ij}$ is the square of the Euclidean distance between the two terms $(i,j)$ [6].

**Application side**

**The second topic:**

**2.1: Introduction**

In this chapter, a type of cluster analysis was applied, which is the Hierarchical cluster analysis method, through data that included (17) governorates of Iraq, Except for the Kurdistan region of Iraq. It included four clusters until they were merged into two clusters according to the closest to the statistical characteristics.

**1-Search variables:**

The data was collected by the Agency of the Ministry of Interior for Police Affairs / General Directorate of Narcotics and Psychotropic Substances Affairs / Directorate of Criminal and Movements / Criminal Statistics Department, and the Hierarchical cluster

analysis method was applied to the variables taken in the study, and these variables were obtained from several primary, including:

1- Governorates
2- Gender (male and female)
3- Age
4- Employee
5- Winner
6- Military
7- Student

The results were interpreted while ensuring the validity of the Hierarchical cluster analysis and the stability of the results. The evaluation of stability is done by using the aggregation method on the same data and testing whether they give the same results in Hierarchical groups as using the distance.

**Table (2.1) Cluster Member ship**

Cluster Membership

| Case | 4 Clusters | 3 Clusters | 2 Clusters |
|---|---|---|---|
| 1: private investigations | 1 | 1 | 1 |
| 2: Baghdad Karkh | 2 | 1 | 1 |
| 3: Baghdad Rusafa | 1 | 1 | 1 |
| 4: Almuthanaa | 1 | 1 | 1 |
| 5: karbala' almuqadasa | 1 | 1 | 1 |
| 6: Babil | 2 | 1 | 1 |
| 7: Alnajaf alasharaf | 1 | 1 | 1 |
| 8: Dhi Qar | 1 | 1 | 1 |
| 9: Dyalaa | 3 | 2 | 2 |
| 10: Albasra | 4 | 3 | 1 |
| 11: Salah aldiyn | 2 | 1 | 1 |
| 12: Maysan | 2 | 1 | 1 |
| 13: Wasit | 4 | 3 | 1 |
| 14: Anbar | 2 | 1 | 1 |
| 15: Aldiywany | 4 | 3 | 1 |
| 16: karkuk | 2 | 1 | 1 |
| 17: Ninawaa | 1 | 1 | 1 |

It was shown in the table (2.1) above in the first stage that: private investigations Baghdad Rusafa ,Al Muthanna, karbala' almuqadasa, Najaf, Ashraf, Dhi Qar, and Ninawaa are within the first cluster, while Baghdad is Karkh, Babil, Salah al-Din, Maysan and Kirkuk within the

second cluster, Dyalaa is within the third cluster, Basra, Wasit and Al-Diwani It is among the fourth cluster .

In the second phase, the governorates that were within the second cluster became part of the first cluster. Which was within the third cluster became within the second cluster. As for the governorates that were within the fourth cluster, they became within the third cluster. As for the last stage, the clusters that were included within the third cluster became within the first cluster. Based on the closeness of the statistical characteristics to each other .

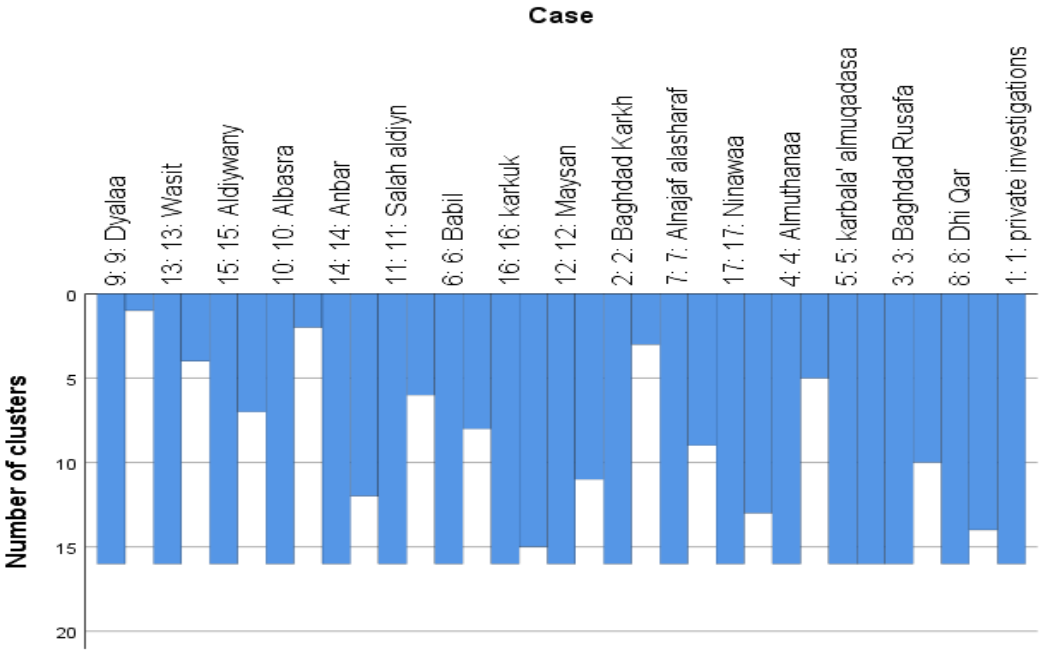


**Fig: (2.1) Alalwah Aljalidia**

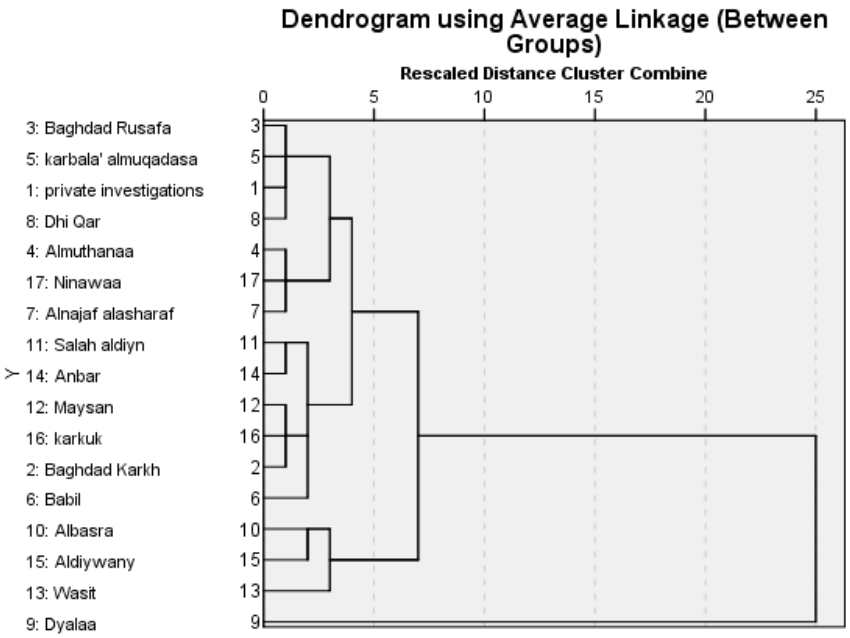


**Fig: (2.2) Hierarchical Tree**

**The third topic**

**3.1: -Discussions:**

1- The first cluster includes several provinces, including private investigations Baghdad Rusafa ,Al Muthanna, karbala' almuqadasa, Najaf, Ashraf, Dhi Qar, and Ninawaa .
1- The second cluster includes several provinces including Baghdad is Karkh, Babil, Salah al-Din, Maysan and Kirkuk.
2- The third cluster includes several provinces including Dyalaa .
3- The fourth cluster includes several provinces including Basra, Wasit and Al-Diwani .
4- All clusters were collected in the first cluster because they have the same statistical characteristics .
5- Diyala Governorate is the only governorate that joined the second cluster.

**3.2: Conclusions:-**

1- Using non-Hierarchical Cluster Analysis and comparing results with Hierarchical Cluster Analysis.
2- Use logistic analysis to analyze data .

**أولاً : المَصادر باللغة العَربية**

**References in Arabic Language**

1. ابراهيم ، عمر سالم ( 2016 ) : " استخدام المؤشرات الصحية لعام 2010 لتصنيف محافظات العراق باستخدام التحليل العنقودي " ، مجلة جامعة تكريت للعلوم الصرفة ، المجلد 21 ، العدد 4 .
2. التحليل العنقودي " مجلة أحمد ، طالب ، " تصنيف المحافظات السورية بحسب الاتفاق الاستهلاكي للأسرة باستخدام جامعة تشرين للبحوث والدراسات العلمية ، سلسلة العلوم الاقتصادية والقانونية ، المجلد 37 ، العدد 2 .
3. أحمد أبو فايد (2016):"التحليل العاملي مفهومه ، أهدافه ، شروطه، أنواعه، مثال تطبيقي لكيفية استخراج التحليل العاملي بنظام SPSS " ، جامعة الازهر –غزة
4. البرزنجي، نظيرة صديق كريم (2002):"دراسة احصائية لتحليل التركيب الكيميائي لبعض الصخور الكاربوناتية في كردستان العراق" رسالة ماجستير، كلية الإدارة والاقتصاد جامعة صلاح الدين.

**ثانيا : المراجع الاجنبية**

5. Aitken, On interpolation by iteration of proportional parts, without use of differences, Proc. Edinburgh. Math . Soc, Series 2, Vol 3, 1932, 56.
6. Alvin, C. Reneher, (2002) “Methods of Multivariate Data Analysis” , A JOHN WILEY & SONS, INC. PUBLICATION, Untied States .
7. Anderson T. W. (1984). An introduction to Multivariate Analysis (22nd ed) . new york John Wiley 1984.
8. Andrée matteaccioli philippe aydalot pionnier de l'économie territoriale l'harmattan 2004 paris France
9. B. Cockburn, G. E. Karniadakis, and C.-W. Shu (eds.), Discontinuous Galerkin methods: theory, computation applications, Lecture Notes in Computational Science and Engi Neering, vol. 11, New York, Springer - Verlag, 2000. MR1842160 (2002b ) .